

SYSTEM FOR AN ADAPTIVE EXCITATION PATTERN FOR SPEECH CODING

INVENTOR

Yang Gao

BACKGROUND OF THE INVENTION

1. Cross Reference to Related Applications.

The present application claims the benefit of U.S. Provisional Application No. 60/233,042, filed September 15, 2000, which is incorporated by reference herein.

The following co-pending and commonly assigned U.S. patent applications were filed on the same day as the above-referenced Provisional Application. All of these applications relate to and further describe other aspects of the embodiments disclosed in this application and are incorporated by reference in their entirety.

United States Patent Application Serial Number _____, "SELECTABLE MODE VOCODER SYSTEM," Attorney Reference Number: 98RSS365CIP (10508.4), filed on September 15, 2000, and is now United States Patent Number _____.

United States Patent Application Serial Number _____, "INJECTING HIGH FREQUENCY NOISE INTO PULSE EXCITATION FOR LOW BIT RATE CELP," Attorney Reference Number: 00CXT0065D (10508.5), filed on September 15, 2000, and is now United States Patent Number _____.

United States Patent Application Serial Number _____, "SHORT TERM ENHANCEMENT IN CELP SPEECH CODING," Attorney Reference Number: 00CXT0666N (10508.6), filed on September 15, 2000, and is now United States Patent Number _____.

United States Patent Application Serial Number _____, "SYSTEM OF DYNAMIC PULSE POSITION TRACKS FOR PULSE-LIKE EXCITATION IN SPEECH CODING," Attorney Reference Number: 00CXT0573N (10508.7), filed on September 15, 2000, and is now United States Patent Number _____.

United States Patent Application Serial Number _____, "SPEECH CODING SYSTEM WITH TIME-DOMAIN NOISE ATTENUATION," Attorney Reference Number: 00CXT0554N (10508.8), filed on September 15, 2000, and is now United States Patent Number _____.

5 United States Patent Application Serial Number _____, "SYSTEM FOR ENCODING SPEECH INFORMATION USING AN ADAPTIVE CODEBOOK WITH DIFFERENT RESOLUTION LEVELS," Attorney Reference Number: 00CXT0670N (10508.13), filed on September 15, 2000, and is now United States Patent Number _____.

10 United States Patent Application Serial Number _____, "CODEBOOK TABLES FOR ENCODING AND DECODING," Attorney Reference Number: 00CXT0669N (10508.14), filed on September 15, 2000, and is now United States Patent Number _____.

15 United States Patent Application Serial Number _____, "BIT STREAM PROTOCOL FOR TRANSMISSION OF ENCODED VOICE SIGNALS," Attorney Reference Number: 00CXT0668N (10508.15), filed on September 15, 2000, and is now United States Patent Number _____.

20 United States Patent Application Serial Number _____, "SYSTEM FOR FILTERING SPECTRAL CONTENT OF A SIGNAL FOR SPEECH ENCODING," Attorney Reference Number: 00CXT0667N (10508.16), filed on September 15, 2000, and is now United States Patent Number _____.

25 United States Patent Application Serial Number _____, "SYSTEM FOR ENCODING AND DECODING SPEECH SIGNALS," Attorney Reference Number: 00CXT0665N (10508.17), filed on September 15, 2000, and is now United States Patent Number _____.

United States Patent Application Serial Number _____, "SYSTEM FOR SPEECH ENCODING HAVING AN ADAPTIVE FRAME ARRANGEMENT," Attorney Reference Number: 98RSS384CIP (10508.18), filed on September 15, 2000, and is now United States Patent Number _____.

30 United States Patent Application Serial Number _____, "SYSTEM FOR IMPROVED USE OF PITCH ENHANCEMENT WITH

SUBCODEBOOKS," Attorney Reference Number: 00CXT0569N (10508.19), filed on September 15, 2000, and is now United States Patent Number _____.

2. Technical Field.

This invention relates to speech communication systems and, more particularly, to systems for digital speech coding.

3. Related Art.

One prevalent mode of communication is by communication systems that include both wireline and wireless radio systems. Data and voice transmissions within a wireless system occur within a bandwidth of an allowed frequency range. Due to increased wireless communication traffic, reduced bandwidth of transmissions to improve capacity with the system is desirable.

Voice and data are transmitted digitally in wireless telecommunications due to noise immunity, reliability, compactness of equipment, and the ability to implement sophisticated signal processing functions using digital techniques. One form of digital transmission is accomplished using digital speech processing systems. Waveforms representing analog speech signals are sampled and then digitally encoded. The number of bits of the encoded signal can be expressed as a bit rate that specifies the number of bits to describe one second of speech. Over the years, significant variations and enhancements have been applied to waveform matching techniques in an effort to improve the quality of the synthesized speech and increase the speech compression.

A reduction in the quality of the synthesized (or reconstructed) speech may occur with respect to the original speech. This divergence in the quality of the synthesized speech is due in part to the failure to closely replicate perceptual aspects of the original speech with the bits of data available to describe the signal. Poor replication of the perceptual aspects could result in noise, loss of clarity and the failure to capture recognizable characteristics such as tone, pitch and magnitude. These characteristics allow a listener to recognize who the speaker is, as well as providing other perception based features, such as, intelligibility and naturalness of the speech.

Accordingly, there is a need for systems of speech coding that are capable of minimizing the bandwidth of original speech, while providing synthesized speech that closely resembles the original speech and captures the perceptually important features of the speech.

SUMMARY

This invention provides a system for an improved excitation enhancement system that uses short term prediction to enhance the excitation signal. As speech data applications continue to operate in areas having intrinsic bandwidth limitations, the perceptual quality of reproduced speech data in typical speech coding systems suffers. The invention employs short term enhancement to improve perceptual quality in reproduced speech.

Speech coding systems may operate using communication media having limited or constrained bandwidth availability. Any communication media may be employed. Examples of such communication media include, but are not limited to, wireless communication media, wire-based telephonic communication media, fiber-optic communication media, and Ethernet.

Other systems, methods, features and advantages of the invention will be or will become apparent to one with skill in the art upon examination of the following figures and detailed description. It is intended that all such additional systems, methods, features and advantages be included within this description, be within the scope of the invention, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE FIGURES

The components in the figures are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention. Moreover, in the figures, like reference numerals designate corresponding parts throughout the different views.

Fig. 1 is an illustration of a waveform illustrating an exemplary speech signal.

Fig. 2 is a block diagram illustrating one embodiment of a speech excitation enhancement system.

Fig. 3 is a block diagram illustrating one embodiment of a speech codec that employs excitation enhancement.

Fig. 4 is a block diagram illustrating another embodiment of a speech codec that employs excitation enhancement.

Fig. 5 is a block diagram illustrating one embodiment of an integrated speech codec that employs excitation enhancement.

Fig. 6 is a diagram illustrating a speech sub-frame depicting excitation enhancement.

Fig. 7 is a functional block diagram illustrating an embodiment of this invention that generates short term enhancement.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

A system is provided that utilizes short term enhancement to enhance coded data that, when decoded, produces a synthesized speech signal that resembles an original speech sample. The system is typically used to enhance speech signals transmitted via a wireless radio telecommunications network. Mobile cellular standards, such as the Adaptive Multi-Rate (AMR) and Selectable Mode Vocoder (SMV) standards, define digital transmission in wireless radio telecommunications. An SMV system is utilized to describe the invention. However, those skilled in the art will appreciate that other systems could be used with the invention.

In Fig. 1, speech coding circuitry (also described in Fig. 2) utilizes prediction to separate a redundant part of a speech signal 100 from an excitation part of the signal 100. The redundant part of the speech signal 100 is an approximately periodic part of the speech signal 100 and the excitation part of the signal describes variations in the speech signal 100. The excitation part of the signal typically may be coded by an encoder and transmitted to a decoder to be converted into synthesized speech (the encoder and decoder are described in Fig. 3). The signals may be coded using a linear predictive coding (LPC) filter. A frame-based algorithm stores sampled input speech signals into blocks of samples called frames 110. An exemplary SMV system operates at a frame size of twenty milliseconds (ms) or one hundred sixty samples per frame. Other sized frames may be used. For signal processing purposes, the frames 110 may be divided into sub-frames 120 that are typically forty samples in size.

Short term enhancement may be used to enhance the excitation signal per sub-frame 120. Short term enhancement utilizes pitch lag information to enhance the excitation signal. Pitch 130 is the approximately periodic part of the speech signal 100, and lag is a measure of the pitch delay in samples. The general shape of the speech signal 100 evolves relatively slowly as a function of time, facilitating pitch prediction and interpolation. By determining information of lag and gain of a sample from a past sub-frame, the information can be scaled and added to a current sub-frame 140 to enhance the limited amount of data generally used to describe the signal for the current sub-frame 140. Thus, a first approximation of the excitation for peak P1 in the current sub-frame 140 is advantageously determined using a scaled segment of the previously sampled value for peak P2. Short term enhancement, further described below with regard to FIG. 6, samples signals within the pitch 130 of a previous sub-frame to approximate corresponding excitation signals in the current sub-frame 140.

Fig. 2 shows a system diagram illustrating one embodiment of an excitation enhancement system 200. The excitation enhancement system 200 may include, among other things, speech enhancement processing circuitry 210, speech coding circuitry 212, long term enhancement circuitry 214, short term enhancement circuitry 216, and speech processing circuitry 218. The speech coding circuitry 212 can include fixed and adaptive codebooks as are known in the art. The speech excitation enhancement system 200 operates on non-enhanced excitation 220 and generates enhanced excitation 230. The speech excitation enhancement system 200 is implemented, for example, on one or more integrated circuits (IC), digital signal processors (DSP) or general processors.

Fig. 3 shows exemplary speech coding circuitry (e.g., speech coding circuitry 212 from FIG. 2) that utilizes enhancement coding 322 at the encoder 320 to perform short term excitation enhancement and long term pitch prediction. A system diagram 300 illustrates one embodiment of a speech codec (e.g., IC with encoder/decoder) that employs speech enhancement in accordance with the invention. A speech encoder 320 of the speech codec 300 performs enhancement coding 322. The enhancement coding 322 is performed using both long term enhancement circuitry 324 and short

term enhancement circuitry 326. The enhancement coding 322 generates prediction and enhancement within the speech sub-frame 120.

The speech encoder 320 of the speech codec 300 also may perform main pulse coding 328 of the speech signal 100 including both sign coding 330 and location coding 332 within the speech sub-frame 120, Fig. 1. Speech processing circuitry 334 also is employed within the speech encoder 320 of the speech codec 300 to assist in speech processing using methods known to those having skill in the art to operate on and perform manipulation of speech data. The speech data, after having been processed, at least to some extent by the speech encoder 320 of the speech codec 300 is transmitted via a communication link 340 to a speech decoder 350 of the speech codec 300. The communication link 340 may be any communication media capable of transmitting voice data, including but not limited to, wireless communication media, wire-based telephonic communication media, fiber-optic communication media, and Ethernet.

The speech decoder 350 of the speech codec 300 may include, among other things, excitation reconstruction circuitry 352, post perceptual compensation circuitry 354, and speech reconstruction circuitry 356. In certain embodiments, the transmit speech processing circuitry 334 and the receiver speech processing circuitry 356 operate cooperatively on the speech data within the entirety of the speech codec 300. Alternatively, the transmit speech processing circuitry 334 and the receiver speech processing circuitry 356 may operate independently on the speech data, each serving individual speech processing functions in the speech encoder 320 and the speech decoder 350, respectively.

The speech processing circuitry 334 and 356 and the main pulse coding circuitry 328 may include, but are not limited to, circuitry and associated algorithms known to those of skill in the art of speech coding. Examples of such main pulse coding circuitry 328 include Code-Excited Linear Prediction (CELP), eXtended CELP (eX-CELP), algebraic CELP and pulse-like excitation. An example of an eX-CELP based speech coder system is described in commonly assigned U.S. Patent App., "SYSTEM OF ENCODING AND DECODING SPEECH SIGNALS," by Yang

Gao, Adil Beyassine, Jes Thyssen, Eyal Shlomot and Huan-Yu Su, previously incorporated by reference.

Fig. 4 illustrates a system diagram of another embodiment of a speech codec 400 that employs excitation enhancement at the speech decoder 450 in accordance with the preferred embodiments. Because the excitation enhancement is performed using data from past sub-frames 120, Fig. 1, the enhancement is accomplished without increasing bandwidth. The speech encoder 410 of the speech codec 400 performs main pulse coding 420 of the speech signal 100 including both sign coding 422 and location coding 424 within the speech sub-frame 120. Speech and excitation processing circuitry 430 also may be employed within the speech encoder 410 of the speech codec 400 to assist in speech processing using methods known to those having skill in the art to operate on and perform manipulation of speech data, examples of which have been previously identified.

The speech data, after having been processed, at least to some extent by the speech encoder 410 of the speech codec 400 may be transmitted via a communication link 440 to a speech decoder 450 of the speech codec 400. The speech decoder 450 of the codec 400 performs excitation enhancement coding 460. The enhancement coding 460 may be performed using both long term enhancement circuitry 462 and short term enhancement circuitry 464. In other embodiments, only short term enhancement is performed. The enhancement coding 460 generates prediction and enhancement within the speech sub-frame 120. The speech decoder 450 of the speech codec 400 may also contain speech reproduction circuitry 470, post perceptual compensation circuitry 480, and excitation reconstruction circuitry 490.

Fig. 5 is a system diagram that illustrates another embodiment of an integrated speech codec 500 that employs speech and excitation enhancement. The integrated speech codec 500 may contain, among other things, a speech encoder 510 that communicates with a speech decoder 520 via a low bit rate communication link 530. The low bit rate communication link 530 may be any communication media capable of transmitting voice data, including but not limited to, wireless communication media, wire-based telephonic communication media, fiber-optic communication media, and Ethernet.

Excitation enhancement coding 540 is performed in the integrated speech codec 500. The enhancement coding 540 may be performed using, among other things, both long term enhancement circuitry 542 and short term enhancement circuitry 544. The long term enhancement circuitry 542 and the short term enhancement circuitry 544 operate cooperatively in certain embodiments, and independently in other embodiments. As shown, the long term enhancement circuitry 542 and short term enhancement circuitry 544 may be arranged within the entirety of the integrated speech codec 500. Depending on the specific application at hand, a user can select to place the long term enhancement circuitry 542 and short term enhancement circuitry 544 in only one or both of the speech encoder 510 and the speech decoder 520. Various embodiments are envisioned, without departing from the scope and spirit of the invention, to place various amounts of the long term enhancement circuitry 542 and the short term enhancement circuitry 544 in the speech encoder 510 and the speech decoder 520. For example, a predetermined portion of the short term enhancement circuitry 544 may be placed in the speech encoder 510 and the remaining portion of the short term enhancement circuitry 544 may be placed in the speech decoder 520.

Figs. 1 and 6 illustrate short term enhancement of the invention. Short term enhancement uses the previous excitation signal to enhance the excitation signal of the current sub-frame 140. The past excitation, weighted by a current weighting filter, may be used to estimate correlation peaks at a distance within the current sub-frame 140. Those skilled in the art will appreciate that an algorithm, similar to that used for long term prediction of pitch lag, can be used to estimate short term correlation of the speech signal 100. In one embodiment, to evaluate short term correlation of the speech signal 100, typically less than five peaks and gains per sub-frame 120 are determined from the past excitation. Those skilled in the art will appreciate that more or less correlation peaks and gains can be determined, depending on the application.

Fig. 6 illustrates a diagram of two pulses I3 and I4 shown at distances R1 and R2 from pulse I2, which correlate to peaks P3, P4 and P2, respectively on Fig. 1. I2 indicates the main pulse, I3 and I4 indicate pulses generated by short term enhancement and Pitch indicates a pulse generated by long term enhancement or short

term enhancement where the true pitch lag is incorrectly determined. The excitation pattern $P(n)$ is constructed as $P(n) = C \sum_i G_i \cdot \delta(n - Ti) + \delta(n)$, where G_i is the gain and T_i is the distance for the i th peak. Regarding Fig. 6, T_0 could equal R1, T_1 could equal R2 and T_N could equal the distance from the main pulse I2 to Pitch. G_0 , G_1 and G_N can correspond to the magnitudes of I3, I4 and Pitch respectively. The gains G_i and the distance T_i may be determined using methods known to those skilled in the art of speech processing. Gains and distances can be calculated, for example, by maximizing correlations of past synthesized signals in a weighted speech domain. The value C is a coefficient typically between 0 and 0.5, and may be a constant or an adaptive value related to the stability of the speech signal. $P(n)$ accounts in part for the fact that the excitation pattern may cover a long term correlation in which the true pitch lag is shorter than the sub-frame size, while the detected pitch lag may be double or triple the true pitch lag.

Fig. 7 is a functional block diagram illustrating an embodiment that generates long term and short term excitation enhancement. In a block 710, a speech signal 100 is processed. In a block, 720, an excitation is coded. In block 730, long term enhancement is performed, and in a block 740, short term enhancement is performed. Additional pulses to the current excitation, as determined by the short term enhancements can be added to the excitation by performing a convolution operation of the excitation pattern $P(n)$ with excitation signals, for example, from a fixed codebook of the speech coding circuitry 512, as known to those of skill in the art. In a block 750, the speech data information is transmitted via a communication link. In a block 760, the speech signal is reconstructed/synthesized.

While various embodiments of the invention have been described, it will be apparent to those of ordinary skill in the art that many more embodiments and implementations are possible that are within the scope of this invention. Accordingly, the invention is not to be restricted except in light of the attached claims and their equivalents.